Large Scale Repository Auditing to ISO 16363

Eloy Rodrigues, University of Minho, eloy@sdum.uminho.pt
Miguel Ferreira, KEEP SOLUTIONS, mferreira@keep.pt
José Carvalho, University of Minho, jcarvalho@sdum.uminho.pt
Pedro Príncipe, University of Minho, pedroprincipe@sdum.uminho.pt
Luís Faria, KEEP SOLUTIONS, lfaria@keep.pt
Hélder Silva, KEEP SOLUTIONS, hsilva@keep.pt
João Moreira, FCT/FCCN, jmm@fccn.pt

Abstract

This paper describes an audit process carried out on 26 digital repositories according to the recently approved standard ISO16363 (Audit and Certification of Trustworthy Digital Repositories). The 26 repositories share a common infrastructure hosted by RCAAP (Open Access Scientific Repository from Portugal), a free hosting service provided to research institutions in Portugal. It addresses the process and the strategic alignment with the project objectives integrated with other developments related to digital preservation of institutional repositories. This work presents the first results of the analysis of the three topics: Organizational Infrastructure, Digital Object Management and Infrastructure and Security Risk Management.

Keywords: digital repository, services, audit, digital preservation, ISO16363,

Audience

Repository managers and decision-makers.

Background

This work included in this presentation covers the following topics:

- Digital preservation tools, services & infrastructure
- Requirements for holding restricted or classified data in repositories
- Infrastructure to accommodate national and international mandates for data management and open access
- Positioning repositories closer to (local, consortia, or cloud-based) cyber infrastructure for data processing

Presentation content

RCAAP[1] is a national-wide project initially funded by UMIC – Knowledge Society Agency, ran by FCT/FCCN, and scientifically supported by the University of Minho that aims at increasing the visibility, accessibility and dissemination of Portuguese research and integrating Portugal in a wide range of international initiatives in the domain of Open Access and infrastructures to support e-science.

On early stages of RCAAP, the project was mostly focused on promoting the creation and development of repositories, by offering repository hosting services, training and advocacy directed to librarians, repository managers and top managers. Additionally, effort was invested on the creation of the RCAAP portal, a search service built to collect, aggregate and index open access scientific results from all the Portuguese institutional repositories.

Over the last years, the RCAAP Project extended its scope building new value-added services for Portuguese scientific community like a Scientific Journal Hosting Service, a Centralized Usage Statistics System for repositories, a Scientific Data repository and, more recently, digital preservation services.

The repository hosting service is a free service that can be used by any scientific or higher education institution that wishes to have its own repository. This service includes not only hosting, but also customization of the repository according to the requirements of the institution, branding, custom configurations and other parameters to adapt the repository to the organizational structure and policies of the adherent institution. The service is supported by DSpace version 1.8.2 and includes a set of custom made add-ons that bring value-added to the service it self.

At the present time, RCAAP's repository hosting service holds 26 repositories from an assortment of institutions ranging from higher education, research institutes, hospitals and laboratories. All repositories share a common infrastructure located and supported centrally by FCT/FCCN, while running operations are managed by the adherent institutions.

Aware of the importance that quality and trustworthiness play as a strategic tool for the national and international scientific infrastructure, the RCAAP project engaged on an audit process across the whole set of repositories centrally hosted.

This work aimed at assessing the level of compliance of the 26 repositories according to the ISO standard 16363:2012 - Audit and certification of trustworthy digital repositories.
The main purpose of this standard is to define a recommended practice on which to base an audit and certification process for assessing the trustworthiness of digital repositories.

The document is meant primarily for those responsible for auditing digital repositories but also for those who are responsible for digital repositories seeking objective

[1] *Repositório Científico de Acesso Aberto de Portugal* or Open Access Scientific Repository from Portugal

measurement of the trustworthiness of their repository. The standard includes normative metrics against which a digital repository may be assessed. These sections are grouped as follows:

1. Organizational Infrastructure;
2. Digital Object Management;
3. Infrastructure and Security Risk Management.

The requirements of the first section were applied to each of the adherent institutions, while the other two were to be fulfilled by the provider of the hosting service.

The auditing process was carried out by a team of 3 auditors from an independent company with high expertise in digital repositories (management and development), digital preservation and consultancy. The auditing process is composed of 4 stages:

1. **Diagnosis** – Repository managers were questioned about their level of compliance on each of requirements that compose the standard. The maturity levels[2] used in the assessment work were based on the ECM3 maturity model[3]. The repository managers were instigated to provide material evidences to support their level of maturity on each of the requirements. In the cases where evidences were not supplied, the lowest level of maturity was assumed.
2. **Action plan** – After the diagnosis stage, the audit team devised a detailed action plan for each of the adherent institutions containing recommendations on how to increase their level of maturity to a level 4 - operational.
3. **Implementation** – Repository managers were given a period of 9 months to implement the action plan.
4. **Final audit** – Finally, a more in-depth audit is expected to take place on each of the repositories to assess their final level of maturity after implementing the recommendations included in the action plan.

Conclusions

The auditing process is currently on stage 3 – implementation - however, some conclusions can already be made.

After the diagnosis stage, we were able to conclude that the average maturity level of all 26 repositories on the standard group of requirements “Organizational Infrastructure” was 2.1 on a scale of 1 to 5. The perceived maturity, i.e. the maturity indicated by repository manager without any concern for gathering evidences of compliance was 2.7.

On the same note, the most mature repository scored 3.0 while the least mature repository scored 1.1.

[2] 1 – Unmanaged, 2 – Incipient, 3 – Formative, 4 – Operational, 5 – Proactive.

[3] http://ecmmaturity.files.wordpress.com/2009/02/ecm3-v2_0.pdf

On the “Digital Object Management” group of requirements, the hosting service provider showed a self-perceptional maturity level of 3.2. Based on the evidences provided, the audit team scored its ability of manage digital objects and associated risks at 2.8.

On the “Infrastructure and Security Risk Management” front, the hosting service showed a self-assessment score of 3.1, while the audit team rank it at 2.6.

One of the difficulties that all the repositories faced was providing evidences that their financial practice is sustainable. Repositories are often managed and operated by library staff embedded on larger institutions such as hospitals or universities. Libraries do not have financial independence nor they care about business models to support their operational work. If one ignores the 3 requirements related to financial sustainability of the repository, the average maturity level of the 26 repositories increases to 2.2.

As stated before, the auditing process is currently on the Implementation stage. The next step consists of wait for the repositories to conclude the implementation stage, scheduled to finish in September 2014. Afterwards, an in-depth audit will be made in order to determine the new maturity level of the audited repositories and supporting infrastructure.